Mixed-State CONDENSATION pour suivi et ré-identification simultanés dans des réseaux de caméras à champs de vue disjoints

Boris Meden, Patrick Sayd, Frédéric Lerasle

Mixed-State CONDENSATION pour suivi et ré-identification simultanés dans des réseaux de caméras à champs de vue disjoints

Boris Meden[1] Patrick Sayd[1] Frédéric Lerasle[2 3]

[1] CEA, LIST, Laboratoire Vision et Ingénierie des Contenus, BP 94, F-91191 Gif-sur-Yvette, France,
[2] CNRS ; LAAS ; 7 avenue du Colonel Roche, F-31077 Toulouse Cedex 4, France,
[3] Université de Toulouse ; UPS, INSA, INP, ISAE ; UT1, UTM, LAAS ; F-31077 Toulouse Cedex 4, France

{boris.meden, patrick.sayd}@cea.fr, lerasle@laas.fr

Résumé

Cet article présente une nouvelle approche pour le suivi de personnes par réseau de caméras à champs de vue disjoints. Le problème du suivi dans l'image est traité par des filtres à particules distribués utilisant un modèle de couleurs hiérarchiques. La nouveauté de notre approche réside dans l'insertion d'une base de personnes déjà rencontrées dans le réseau, dans le formalisme du filtre à particule. Ce faisant, les filtres ne réalisent plus seulement une estimation de position dans l'image mais aussi établissent une identité potentielle pour les cibles, relativement à la base de personnes. Ainsi nous envisageons la ré-identification en ligne de personnes pour introduire de la continuité et pouvoir suivre les cibles dans un réseau à champs de vue disjoints. Aucune calibration n'est requise. Nous évaluons notre approche sur un réseau de 5 caméras à champs de vue disjoints et un ensemble de 16 personnes.

Mots Clef

Ré-identification, suivi, réseau de caméras, champs de vues disjoints, filtrage particulaire.

Abstract

This article presents a novel approach to person tracking within large-scale environments monitored by non-overlapping field-of-view camera networks. We address the image-based tracking problem with distributed particle filters using a hierarchical color model. The novelty of our approach resides in the embedding of an already-seen-people database in the particle filter framework. Doing so, the filter performs not only image position estimation but also does establish identity probabilities for the current targets in the network. Thus we use online person re-identification as a way to introduce continuity to track people in disjoint camera networks. No calibration stage is required. We demonstrate the performances of our approach on a network of 5 disjoint cameras and a 16-person database.

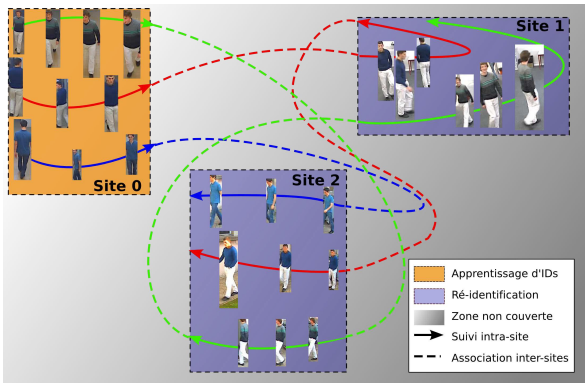

FIGURE 1 – Illustration du processus de ré-identification sur un réseau de caméras en trois sites distincts ne présentant aucuns champs de vue communs. Les cibles présentes dans le réseau sont référencées sur le premier site, puis ré-identifiées dans les suivants. Ceci permet d'inférer des trajectoires (traits pointillés) au niveau du réseau et d'apporter une continuité aux suivis intra-sites.

Keywords

Re-identification, tracking, camera network, non-overlapping fields of view, particle filtering.

1 Introduction

Le problème de l'estimation de la trajectoire d'un objet se déplaçant dans une région d'intérêt, abrégé tracking, est un sujet de recherche majeur en vision par ordinateur (voir une étude complète dans [3]). Le problème devient d'autant plus difficile lorsque des pistes multiples (MOT Multiple Object Tracking) sont suivies conjointement, avec pour but le maintien des identités de cibles. MOT a été traité par des approches supervisées [14], mais aussi avec des filtres à particules distribués [13] [1]. Cependant, il n'est généralement pas envisageable de couvrir complètement de larges zones avec des caméras à champs de vue communs pour des raisons économiques et/ou de temps de calcul. Ainsi, pour des scenarii réalistiques, le système se doit de pou-

voir gérer des caméras multiples ne présentant pas de recouvrements dans leurs champs de vue. Au-delà du problème de suivi dans chaque caméra, la principale difficulté réside dans les transitions entre caméras et le problème de maintien d'identités de pistes au niveau du réseau.

Ce saut entre caméras, connu comme le problème de ré-identification, peut être divisé en deux sous-problèmes. Il trouve ses solutions d'une part dans la capacité du descripteur à être robuste au changement de caméras (en terme de point de vue, tout comme de réponse colorimétrique du capteur), et d'autre part, dans la stratégie d'appariement de descripteurs d'une caméra à l'autre. Comme la plupart des approches récentes, Gray *et al.* [5] se concentrent sur le descripteur et visent le meilleur taux de ré-identification image à image. Ils proposent par ailleurs la base de données VIPeR, présentant des collections d'images de personnes sous deux points de vue différents. Prosser *et al.* adoptent une démarche similaire dans [12]. Plutôt que de choisir un descripteur de cible particulier, ces travaux ont recours à un méta-algorithme comme le boosting dans [5], pour mettre en avant les caractéristiques discriminantes entre des paires correctes d'images issues d'un ensemble apprentissage. Les limitations de ces approches sont le grand nombre d'images labelisées nécessaires à l'apprentissage de la frontière de décision. D'autres travaux, travaillant aussi au niveau descripteur, cherche à projeter les descripteurs issus de deux capteurs différents sur le même sous-espace. Ils mettent l'accent sur la réponse colorimétrique et cherchent à la calibrer. Ainsi, Javed *et al.* dans [7] estiment une fonction de transfert colorimétrique sur une collection de cibles d'entrainement vues dans le réseau. Bowden *et al.* poursuivent cet apprentissage dans [4] en l'estimant de façon incrémentale. Là encore, les limitations se trouvent dans la phase nécessaire d'entrainement, qui en outre biaise la fonction de transfert sur les couleurs des cibles sur lesquelles elle est entrainée.

Les approches évoquées jusqu'à maintenant considèrent toutes une stratégie d'appariement simple, image à image. Cong *et al.* dans [2] proposent un processus d'appariement plus évolué en considérant des collections d'images clés représentant le passage d'une cible dans une caméra. Pour toute paire de séquences, ils font une analyse spectrale du graphe laplacien de la matrice de similarité entre les séquences. Cette méthode de réduction de dimension permet de décider simplement si les images clés sont suffisament proches pour correspondre à la même identité vue dans deux caméras ou non. La ré-identification entre caméras est rendue possible par l'utilisation de la normalisation Greyworld sur l'espace RGB, ainsi que cette stratégie élaborée d'appariement. Cependant, la comparaison n'est valable que pour une paire de séquences à la fois. Makris *et al.* dans [9] proposent d'apprendre les transitions spatio-temporelles dans les zones noires du réseau, de manière à pouvoir estimer des temps de parcours de ces zones. Avec le même objectif, Lim *et al.* dans [8] proposent un filtre à particule muni de deux comportements, selon que la personne est visible dans le réseau ou non. Dans le premier cas, ils suivent la cible dans le plan image. Ils font l'hypothèse de disposer d'une carte métrique du bâtiment où le réseau est déployé. Lorsque la cible sort du champ de vue du réseau, ils propagent à vitesse constante un nuage de particules dans les plans du bâtiment, le divisant à chaque intersection. La ré-identification intervient en donnant à une détection l'identité des particules propagées les plus proches. L'approche est limitée par le traitement de cibles multiples.

Dans ce papier, nous envisageons le suivi dans des réseaux à champs disjoints comme une extension du suivi multi-cibles [14] [13]. Pour ce faire, nous proposons d'introduire la ré-identification dans le formalisme de filtrage particulaire. Ainsi nous n'estimons pas seulement une position relative dans une caméra donnée, mais aussi une identité pour la cible courante relativement à une base de personnes présentes dans le réseau, apprise *a priori* (typiquement, les personnes qui ont été suivies une première fois dans une caméra de hall de bâtiment). En outre, nous utilisons des filtres distribués, ce qui induit une faible complexité et rend notre approche extensible à un nombre important de caméras. Cette stratégie peut-être vue comme une comparaison image par image évoluée car le processus de filtrage introduit une temporalité, tout en produisant toujours un résultat de ré-identification à chaque image traitée. Étant donné qu'il n'existe pas de base de données publique traitant de réseaux étendus à champs disjoints, nous avons testé nos algorithmes sur notre propre réseau, composé de 5 caméras déployées dans un couloir de 34m, une salle de réunion et une zone extérieure, avec un total de 16 personnes évoluant dedans.
Dans la suite, la section 2 commence par présenter la manière dont nous apprenons les identités que nous chercherons à suivre et ré-identifier par la suite. La section 3 détaille l'ajout de la ré-identification dans le filtrage particulaire. La section 4 introduit une notion de superviseur, pour coordonner les filtres distribués dans un contexte multi-pistes. Et finalement, la section 5 presente les évaluations de notre approche.

2 Apprentissage des identités à ré-identifier

2.1 Représentation de la cible

Pour éviter les problèmes de calibration géométrique, le suivi est réalisé dans le plan image. Nous utilisons un modèle géométrique rectangulaire pour définir la région d'intérêt de la cible (ROI). Cette ROI est découpée en bandes horizontales régulières et chacune est décrite par sa distribution de couleur. Les histogrammes de couleur ont montré leur robustesse aux changements d'apparences [10] avec leur aspect global. L'ajout de contraintes spatiales dans la signature, au travers du découpage en bandes, localise les couleurs et accroit son pouvoir discriminant. Cette signature a été utilisée avec succès pour du suivi par Pérez

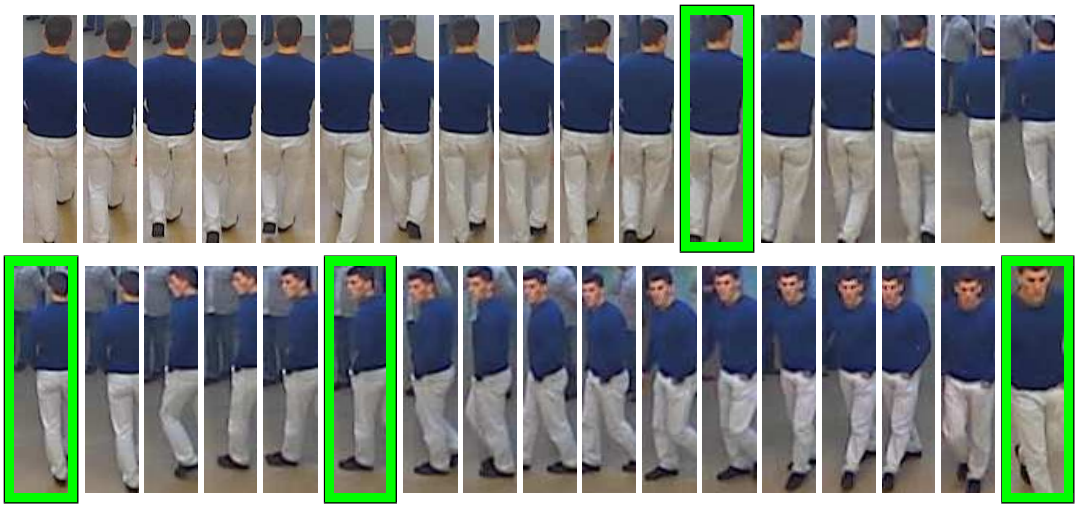

FIGURE 2 – Quelques ROI faisant partie d'une séquence de suivi (ramenées à la même hauteur pour le papier). Les cadres verts mettent en valeurs les quatres images clés retenues par la méthode pour décrire la séquence. Elles capturent les plus grandes variations d'apparence de la cible au cours de la séquence.

et al. dans [11] tout comme pour la ré-identification par Cong *et al.* dans [2]. En outre, ce type de signature (appelé Hand Localized Histogram dans leur papier) a fait partie de la campagne d'évaluations menée par Gray *et al.* dans [5], et fournissait de bons résultats pour la stratégie de comparaison image à image. Nous utilisons des histogrammes couleur dans l'espace couleur RGB, avec une quantification de 8 accumulateurs par canal pour des raisons de temps de calcul dans le suivi. Le nombre de bandes a lui été optimisé en utilisant le mode opératoire de [5]. Après avoir calculé les courbes Cumulative Matching Characteristic (CMC) sur la base VIPeR, pour un nombre de bande variant de 1 à 30, nous avons conservé la meilleure courbe, correspondant à 5 bandes. Associée à un processus de normalisation détaillé en sous-section 3.2, l'utilisation d'accumulateurs larges permet d'absorber en partie le décalage de couleur d'un capteur à l'autre.

2.2 Réduction en images clés

Avant de pouvoir reconnaître une personne, il faut l'avoir vue une première fois. Nous proposons d'apprendre les identités qui vont évoluer dans le réseau dans une première qui sera vue comme le point d'entrée dans le réseau (*e.g.* le hall d'un building) et le Site 0 dans nos expériences (Figure 5). Par la suite, nous allons traiter le réseau comme un système fermé, avec une collection fixe d'identités. Nous commençons donc par exécuter un filtre à particule CONDENSATION [1] sur la séquence d'apprentissage dans la première caméra. Ainsi nous extrayons une vue de la cible à chaque instant. Ensuite, nous réduisons hors ligne cette collection de descripteurs en images clés. Pour sélectionner le bon nombre d'images clés par identité, de manière à conserver le maximum de variabilité dans l'apparence, nous réalisons une analyse spectrale des suivis. Nous approche s'inspire de [2] mais se limite ici à une seule personne. Ainsi, nous focalisons sur les variations dans une séquence de descripteurs. Pour cela, nous construisons la matrice de similarité de chaque séquence de suivi dans la caméra d'apprentissage comme $W_{ij} = \exp(-K \cdot \sum_{k=1}^{N_c} d^2(s_i(k), s_j(k)))$, où $d(.,.)$ est la distance de Bhattacharyya discrète, $s_i(k)$ (*resp.* $s_j(k)$) est la k-ième distribution de couleurs de la cible i (*resp.* j), N_c est le nombre de bandes par cible et K une constante de normalisation. Nous appliquons une méthode de Spectral Clustering à cette matrice de similarité en calculant son Graphe Laplacien non normalisé $\Delta = D - W$ où D la matrice diagonale des sommes des lignes de W : $D_{ii} = \sum_j W_{ij}$. La diagonalisation de Δ renseigne sur les clusters de W. En effet, les valeurs propres présentent un saut lorsque le nombre de clusters est atteint [2]. Pour une seule personne, le saut est généralement peu prononcé, ce qui correspond à une répartition des points en clusters non triviaux. Pour cela, nous seuillons les valeurs propres en ne conservant qu'un pourcentage de la valeur propre la plus élevée (15 % sur les séquences du réseau de test). Nous réalisons ensuite un clustering dans l'espace réduit des k premiers vecteurs propres par k-means, en k classes, k étant le nombre de valeurs propres moins un inférieures au seuil. Nous résumons ainsi des séries d'images par celles qui présentent le plus de variations au niveau de l'apparence. La figure 2 illustre cette sélection. Pour une centaine de boites de suivi, nous extrayons entre 4 et 10 images clés, selon la séquence. L'avantage de cette méthode est de ne pas fixer *a priori* le nombre d'images clés, mais de l'adapter à la séquence d'imagettes considérée.

1. Pour Conditional Density Propagation.

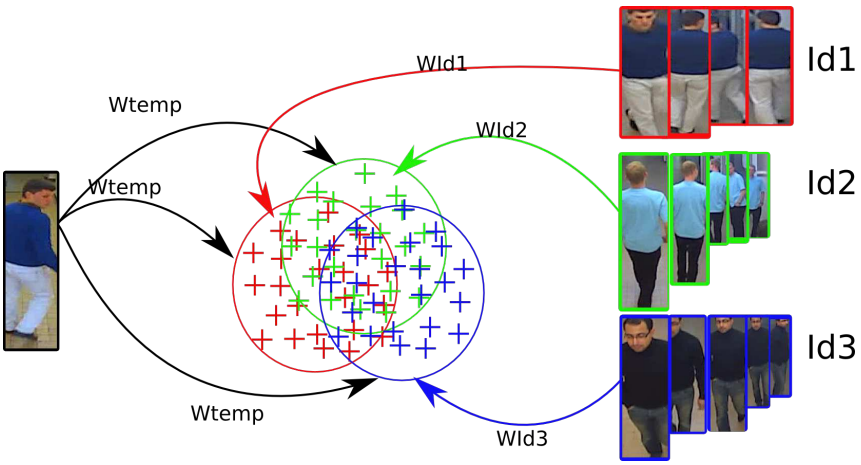

FIGURE 3 – Illustration de notre Mixed State CONDENSATION dans le cas d'une base de cardinalité 3. Le nuage de particules est divisé en trois sous-nuage, identiquement distribués à l'initialisation du filtre. Par la suite, l'identité la plus ressemblante va prendre le dessus grâce à la matrice de transition d'identités et à la vraisemblance combinée. Les particules partagent la même référence temporelle de suivi (à gauche), mais différentes identités dans la base (à droite avec les images clés).

3 Insertion de la ré-identification dans le processus de suivi

3.1 Formalisme de filtrage particulaire

Un processus de suivi de type filtrage bayésien commence par l'initialisation d'une région de référence dans la l'image (il s'agit de notre référence temporelle, figure 3), et procède ensuite à la recherche récursive de régions similaires au sens de la description qui en est faite, dans la suite de la séquence vidéo. Étant donnée la base d'identités, nous avons des descripteurs de référence supplémentaires auxquels se comparer. Pour cela, nous utilisons un filtre à particule de type Mixed State CONDENSATION, introduit dans [6]. Nous cherchons à estimer un vecteur d'état mixte, composé de paramètres continus (la position de la boite dans l'image $\mathbf{x}$), mais aussi un paramètre discret (l'identité de la cible y), dans la boucle de filtrage, soit

$$\mathbf{X} = (\mathbf{x}, y)^{\mathsf{T}},\ \mathbf{x} \in \mathbb{R}^4,\ y \in \{1, \dots, N_{id}\}$$

Dans notre cas, le suivi est réalisé dans le plan image avec un modèle géométrique rectangulaire paramétré par $\mathbf{x} = [x_c, y_c, h_x, r]^{\mathsf{T}}$, où $(x_c, y_c)^{\mathsf{T}}$ sont les coordonnées du centre de la boite, h_x sa demi-largeur, r le ratio hauteur-largeur, N_{id} le cardinal de la base d'identités, et N le nombre de particules. Étant donné ce vecteur d'état étendu, la densité du processus d'échantillonnage à l'image t peut être écrite comme dans [6] :

$$p(\mathbf{X}_t|\mathbf{X}_{t-1}) = T(\mathbf{X}_t, \mathbf{X}_{t-1}) \cdot p(\mathbf{x}_t|\mathbf{x}_{t-1})$$

où $T(\mathbf{X}_t, \mathbf{X}_{t-1})$ est la matrice de probabilité de transitions, appliquée au paramètre discret d'identité, et $p_{ij}(\mathbf{x}_t|\mathbf{x}_{t-1})$ est l'échantillonnage de la loi appliquée à la partie continue de l'état. La matrice de transition $T = [t_{ij}]$ est construite sur l'ensemble des images clés. L'élément t_{ij} est la similarité entre les identités i et j de la base, calculée en utilisant l'équation (1) entre les images clés les plus dissemblables. Dans [6] le paramètre discret est utilisé pour considérer deux modèles de mouvement concurrents, et laisser le filtre décider lequel correspond le mieux à la situation courante. Dans notre cas, ce paramètre référence une identité dans la base de personnes présentes dans le réseau. Le nuage de particules est divisé en sous-nuages, associés chacun à une identité.

3.2 Estimation conjointe de l'identité et de la position

Après l'étape d'échantillonnage, les nouvelles positions de particules sont évaluées. La vraisemblance temporelle de la CONDENSATION $p(\mathbf{Z}_t|\mathbf{x}_t^{(n)})$ est approximée par :

$$w_{Temp}^{(n)}(t) = \exp\{-K \cdot \sum_{j=1}^{N_c} d^2\left(s_t^{(n)}(j), s_{Temp}(j)\right)\},$$
$$\forall\, n = 1, \dots, N$$

où N_c est le nombre de distributions de couleurs par cible, $s_{Temp}(.)$ l'ensemble des distributions de couleurs de la référence temporelle, $s_t^{(n)}(.)$ l'ensemble des distributions de couleur de la particule courante, et N est le nombre de particules.
Le formalisme du Mixed-State CONDENSATION adapté à la ré-identification fournit une vraisemblance additionnelle, pondérant les particules relativement à leur identité de référence, $p(\mathbf{Z}_t|\mathbf{x}_t^{(n)}, y_t^{(n)})$:

$$w_{Id}^{(n)}(t) = \exp\{-K \cdot \min_{i \in N_y} \sum_{j=1}^{N_c} d^2\left(s_t^{(n)}(j), s_{identity}(j, y_t^{(n)}, i)\right)\},$$
$$\forall\, n = 1, \dots, N \qquad (1)$$

où N_y est le cardinal de la classe d'images clés de l'identité $y_t^{(n)}$ ($y_t^{(n)}$ étant l'identité assignée à la n-ième particule au temps t), N_c le nombre de distributions de couleurs par cible, $s_{identity}(., y_t^{(n)}, i)$ est l'ensemble de distributions de couleurs de la i-ième image clé de l'identité $y_t^{(n)}$ dans la base, $s_t^{(n)}(.)$ est l'ensemble de distributions de couleurs de la particule courante, et N est le nombre de particules. La figure 3 résume le principe de ces deux vraisemblances par particule. Chacune est évaluée relativement à la référence temporelle de suivi (w_{Temp}), mais aussi (w_{Id}) relativement à son identité (décrite par une collection d'images clés).

w_{Temp} est évaluée sur des descripteurs issus de la même caméra, à des instants proches, alors que pour w_{Id}, ils proviennent de deux caméras différentes. Les ordres de grandeurs de ces vraisemblances ne sont pas les mêmes. Pour pouvoir toutefois combiner ces vraisemblances, nous les normalisons chacunes sur l'ensemble des particules avant l'étape de ré-échantillonnage du filtre. De la même manière que [7] et [4] proposent l'estimation d'une fonction de transfert colorimétrique inter-caméras, la normalisation que nous introduisons ici nous permet de relever la réponse de la caméra d'apprentissage et de ré-identifier les pistes actives de suivi. Nous évitons une estimation coûteuse et propre à une paire de caméras et nous garantissons ainsi le changement de caméras, connu comme ré-identification. Nous notons w_{Id}^* et w_{Temp}^* les vraisemblances normalisées.

Si w_{Temp}^* est supérieur à un seuil (*i.e.* si la particule est intéressante, sinon conservons cette faible vraisemblance temporelle comme vraisemblance combinée), nous combinons ces deux similarités normalisées pour obtenir l'expression de la vraisemblance de particule qui sera injectée dans l'étape de pondération du filtre à particules :

$$\pi_t^{(n)} = \alpha \cdot w_{Temp}^{*(n)}(t) + (1-\alpha) \cdot w_{Id}^{*(n)}(t),\ \forall\, n = 1, \ldots, N.$$

Ce faisant, nous donnons de l'importance aux particules correctement positionnées, qui possèdent la bonne identité. L'estimation de l'état est ensuite un processus en deux étapes. Nous commençons par calculer le Maximum A Posteriori sur le paramètre discret, *i.e.* l'identité la plus probable à l'instant, soit une ré-identification partielle.

$$\hat{y}_t = \arg\max_j P(y_t = j|\mathbf{Z}_t)$$

$$= \arg\max_j \sum_{n \in \Upsilon_j} \pi_t^{(n)}, \text{ où } \Upsilon_j = \left\{ n | s_t^{(n)} = (\mathbf{x}_t^{(n)}, j) \right\} \quad (2)$$

Ensuite, les composantes continues sont estimées sur le sous-ensemble de particules qui possèdent l'identité la plus vraisemblable.

$$\hat{\mathbf{x}}_t = \sum_{n \in \hat{\Upsilon}} \pi_t^{(n)} \cdot \mathbf{x}_t^{(n)} / \sum_{n \in \hat{\Upsilon}} \pi_t^{(n)}, \text{ où } \hat{\Upsilon} = \{ n | s_t^{(n)} = (\mathbf{x}_t^{(n)}, \hat{y}_t) \} \quad (3)$$

4 La contrainte de non-ubiquité

Notre approche distribuée fournit une stratégie pour la ré-identification. Plutôt que de comparer une image requête à toutes les entrées de la base d'identité, nous laissons notre filtre à particules Mixed-State réaliser l'estimation, permettant ainsi la gestion d'identités proches. Les limites de l'approche se situent dans l'absence d'intéractions entre les filtres. Rien n'empêche le choix d'une même identité au même instant pour deux filtres différents.

Pour éviter cette incohérence, nous ajoutons à l'approche une légère procédure de supervision, qui réunit les probabilités de ré-identification à chaque instant grâce à la caractérisation en ligne des identités, et qui assigne à chaque filtre son identité la plus probable, en tenant compte des autres filtres. Pour un cas de suivi multi-cibles, l'équation 2 est remplacée par l'équation 4.

$$\hat{y}_t(f) = \arg\max_j P(y_t = j|\mathbf{Z}_t, f) = \arg\max_j \sum_{n \in \Upsilon_{j(f)}} \pi_t^{(n)}(f), \quad (4)$$

où $\Upsilon_j(f) = \left\{ n | s_t^{(n)}(f) = (\mathbf{x}_t^{(n)}(f), j) \right\},\ \forall f = 1 \ldots N_{filters}$

où $(s_t^{(n)}(f), \pi_t^{(n)}(f))$ représente la n-ième particule et sa vraisemblance, du f-ième filtre, et $N_{filters}$ est le nombre de filtres actuellement en cours. Quand un filtre se voit affecter une identité, cette identité devient impossible pour les filtres restants. De cette manière, nous empêchons deux filtres de choisir la même identité.

5 Évaluations

5.1 Description du réseau d'évaluation

Nous avons utilisé un réseau de cinq caméras ne présentant aucun recouvrement de champs de vue (Figure 5). La caméra 0 a été utilisée pour réaliser l'apprentissage de la base, composée de 16 personnes (la Figure 4 montre une image clé par identité).

5.2 Efficacité de la ré-identification

Comme nous l'avons expliqué en section 1, notre approche Mixed-State amène une nouvelle stratégie au problème de ré-identification. Pour évaluer cette approche, nous commençons par proposer une comparaison de cette stratégie à celle de l'état de l'art, dans le cas de suivi d'une seule cible. Nous calculons les taux de ré-identification pour les 16 identités de la base, dans chacune des caméras avec une stratégie image par image à chaque instant et avec notre stratégie. Dans les deux cas, nous utilisons le même descripteur (détaillé en section 2) car nous évaluons les stratégies d'appariement. Les initialisations de pistes ont été réalisées à la main. Pour la comparaison image à image, nous utilisons un suivi sans ré-identification, et à chaque nouvelle estimation de position, nous la comparons à chacune des entrées de la base. À cela nous confrontons la stratégie Mixed State. La vérité terrain sur les identités permet d'obtenir une réponse binaire pour chaque stratégie, pour chaque cible et à chaque instant. Pour chaque caméra, nous sommons les résultats obtenus sur les cibles, ce

FIGURE 4 – Notre ensemble de 16 personnes traversant le réseau.

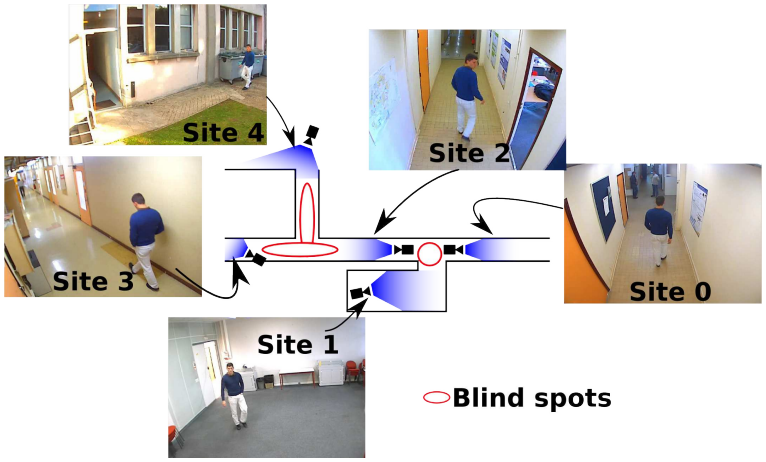

FIGURE 5 – Vue d'ensemble du réseau de test, composé d'un couloir de 34m, une salle de réunion et une zone extérieure.

TABLE 1 – Taux de ré-identification entre caméras pour la stratégie triviale et la stratégie Mixed-State.

	Track then ID	Track + ID
Site #0 to #0	0.96	**0.98**
Site #0 to #1	0.40	**0.46**
Site #0 to #2	0.66	**0.81**
Site #0 to #3	0.65	**0.71**
Site #0 to #4	0.30	**0.34**

qui donne des taux de ré-identification par image. Nous moyennons ensuite ces taux sur la séquence. De plus les taux sont moyennés sur cinq répétitions de chaque suivi pour prendre en compte la nature stochastique du filtrage particulaire. La table 1 résume les résultats obtenus.

Nous observons différents taux de ré-identification dépendant de la caméra considérée. Le site #0 est celui où les identités ont été apprises, donc les descripteurs sont vraiment similaires. Ceci explique le taux approchant les 100%. En revanche, les sites #1 et #4 sont relativement différents en terme de pose de caméra et de couleurs de fond (le site #4 étant de plus en extérieur et présentant de forts changements d'illuminations). Le descripteur choisi n'utilise pas de soustraction de fond, ce qui est une des explications à la chute des taux. De plus, aucune normalisation de l'espace couleur n'est encore utilisée. Cependant, nous constatons pour toutes les caméras, que la stratégie de suivi et ré-identification simultanés est toujours meilleure.

5.3 Ré-identification en configuration multi-cibles

La Figure 6 donne une illustration qualitative du cas d'utilisation type de la contrainte de non-ubiquité : plusieurs cible évoluant simultanément dans le réseau, et éventuellement dans la même caméra. La vérité terrain sur les identités donne les id 1 pour la caméra 3, et 6, 3, 9 et 10 pour la caméra 2. A cet instant, tous les filtres ré-identifient correctement leurs cibles.

6 Conclusion et perspectives

Nous avons proposé une nouvelle approche pour le suivi de personnes dans des réseaux de caméras à champs de vue disjoints, qui ne requiert aucune connaissance *a priori* sur le réseau. Nous voyons la ré-identification de personnes comme un moyen d'introduire de la continuité en ligne entre des séquences de suivi dans différentes caméras. La principale nouveauté de ce papier est l'introduction de cette ré-identification dans le formalisme de filtrage particulaire, au moyen de l'algorithme Mixed-State CONDENSATION. Le but est d'estimer simultanément la position de la cible et son identité dans le réseau. Plutôt que de mettre les efforts sur l'apprentissage d'un descripteur, nous proposons ici une stratégie d'appariement évoluée pour le problème de ré-identification. De plus, cette stratégie s'inscrit directement dans la démarche de suivi. Nous avons montré par une comparaison approfondie, sur les sites de notre réseau, que cette stratégie surpasse la comparaison image à image traditionnelle. En outre, notre approche est théoriquement

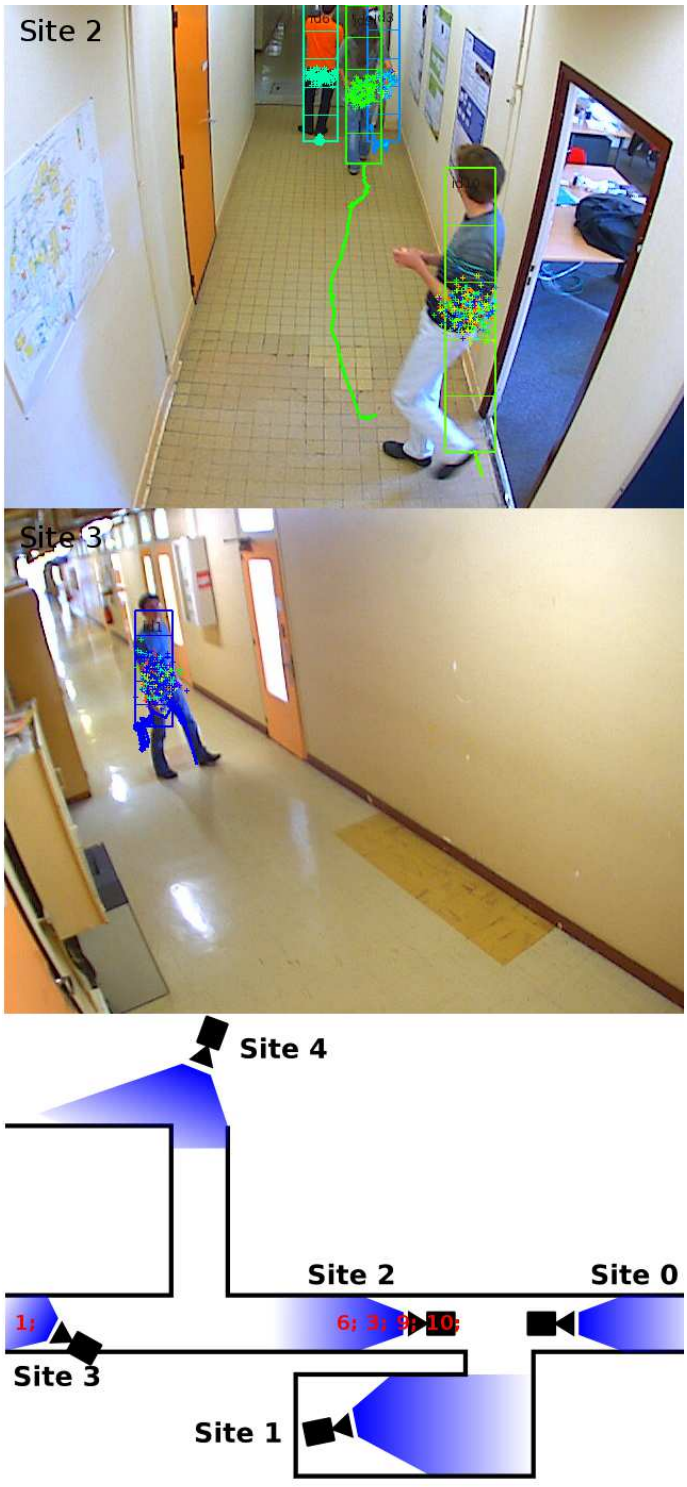

FIGURE 6 – Suivi de cinq cibles dans le réseau : quatre dans le site #2 et une dans le site #3. Les résultats de ré-identification sont reportés sur une carte grossière du réseau.

indépendante du nombre de caméras comme les filtres sont distribués. De plus, nous proposons un moyen de contraindre la non-ubiquité d'identités, pour les cas de suivi multi-cibles.

Les travaux futurs s'intéresseront à une construction et mise à jour en ligne de la base d'identités évoluant dans le réseau. Les intéractions entre filtres au sein de la même caméra pour éviter la dérive de filtres sur la même cible n'ont pas été prises en compte dans ce papier. L'ajout de forces d'intéractions comme proposé dans [13] renforcera le suivi multi-cible mono-caméra. Et finalement, notre approche n'utilise que l'information image. L'ajout de connaissance *a priori* telle que le plan du sol, ou une carte topologique du réseau pour une gestion plus fine des identités, sont autant de compléments qui peuvent venir renforcer la méthode.

Références

[1] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Robust tracking-by-detection using a detector confidence particle filter. In *Proceedings of the International Conference on Computer Vision*, 2010.

[2] D.-N. Truong Cong, L. Khoudour, C. Achard, C. Meurie, and O. Lezoray. People re-identification by spectral classification of silhouettes. *Signal Processing*, 2009.

[3] P.F. Gabriel, J.G. Verly, J.H. Piater, and A. Genon. The state of the art in multiple object tracking under occlusion in video sequences. *Advanced Concepts for Intelligent Vision Systems*, 2003.

[4] A. Gilbert and R. Bowden. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *Proceedings of the European Conference on Computer Vision*, 2006.

[5] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of the European Conference on Computer Vision*, 2008.

[6] M. Isard and A. Blake. A mixed-state condensation tracker with automatic model-switching. In *Proceedings of the International Conference on Computer Vision*, 1998.

[7] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2005.

[8] FL Lim, W. Leoputra, and T. Tan. Non-overlapping distributed tracking system utilizing particle filter. *The Journal of VLSI Signal Processing*, 2007.

[9] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *Proceedings of the International*

Conference on Computer Vision and Pattern Recognition, 2004.

[10] K. Nummiaro, E. Koller-Meier, and L. Van Gool. An adaptive color-based particle filter. *Image and Vision Computing*, 2003.

[11] P. Perez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. In *Proceedings of the* IEEE, 2004.

[12] B. Prosser, W.S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person Re-Identification by Support Vector Ranking. In *Proceedings of the British Machine Vision Conference*, 2010.

[13] Wei Qu, Dan Schonfeld, and Magdi Mohamed. Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *EURASIP J. Appl. Signal Process.*, 2007.

[14] K. Smith, D. Gatica-Perez, and J.M. Odobez. Using particles to track varying numbers of interacting people. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2005.